3

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 January 2005 (13.01.2005)

PCT

(10) International Publication Number
WO 2005/004366 A2

(51) International Patent Classification⁷: H04L

(21) International Application Number: PCT/IL2004/000591

(22) International Filing Date: 1 July 2004 (01.07.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/483,909 2 July 2003 (02.07.2003) US
10/759,091 20 January 2004 (20.01.2004) US

(71) Applicant *(for all designated States except US)*: YISSUM RESEARCH DEVELOPMENT COMPANY OF [IL/IL]; Hi Tech Park, The Edmond J. Safra Campus, The Hebrew University of Jerusalem, Givat Ram, 91 390 Jerusalem (IL).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: KIRKPATRICK, Scott [US/IL]; 15 Neve ShaAnan Street, 93708 Jerusalem (IL). WEINSHALL, Daphna [IL/IL]; 15 Neve ShaAnan Street, 93708 Jerusalem (IL).

(74) Agent: G.E. EHRLICH (1995) LTD.; 11 Menachem Begin Street, 52 521 Ramat Gan (IL).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: IMPRINTING AN IDENTIFICATION CERTIFICATE

(57) Abstract: A system and method for using imprinting as part of a security function that involves a user, for a security and/or identification mechanism. Imprinting is preferably used for cryptographic primitives, for determining a one-way function that operates at least partially according to a characteristic and/or function of the human brain.

IMPRINTING AN IDENTIFICATION CERTIFICATE

FIELD OF THE INVENTION

The present invention relates to the use of human memory as a security and/or identification mechanism, and in particular, the use of imprinting for such a mechanism.

BACKGROUND OF THE INVENTION

Security is increasingly important, such that security and/or identification mechanisms are also very important. Such mechanisms are vulnerable to attack through stealing passwords or codes. One attempt to overcome such vulnerabilities is to use an aspect of the human individual which cannot easily be copied, such as fingerprints or retinal "prints". However, these biometric characteristics may still be copied or forged.

The protocol used to verify a password is quite simple, and usually involves comparing an encrypted version of the password with a stored encrypted copy. The weakness is the difficulty of remembering all the passwords and PINs that modern life requires without writing them all down (unencrypted) and posting them in an obvious place or using easily-guessed personal information. As a result, the apparent security of a password can be illusory. Programs such as L0phtCrack and its commercial derivative, L4 (Password), have shown that many passwords can be guessed by attacks which try long lists of common words, enhanced by random extensions.

The common PIN or password is easy to describe to others. One can easily be impersonated by someone who knows the password, such that it is not very safe from eavesdroppers. Users must also make considerable effort to remember all of the passwords being used. However, the protocol used to verify a password is quite simple. Therefore, solutions to the above problems should also be easy to use, but also safer from impersonation. Verification should also be easy.

Some attempts to overcome these problems have involved maintaining at least some information in the mind of a user. This information may then be used for security and/or identification, optionally through some type of mental function or effort (other than memory alone).

Previous efforts to create better schemes for identifying human individuals to computers have focused on defeating the efforts of an observer or wire-tapping eavesdropper by requiring the human individual to perform mathematical calculations involving a shared secret. Matsumoto's 1991 and 1996 papers, for example, require the user to perform the XOR of a supplied bit string with a memorized bit string, and report the parity of the result to the computer, and/or other calculations. While this may be within the mental arithmetic capabilities of some programmers, it seems too complex for general use. The method requires the use of a shared secret password, in this case a secret bit string, of which the user is completely aware. Hopper and Blum (2000) explore more complicated protocols which, they argue, reduce the complexity of the computation that the human individual must perform.

The literature of psychophysics and cognitive psychology has many studies of "imprinting" phenomena, simple behaviors or recognitions which are quickly learned, and can be retrieved much later with little effort. Obviously the "imprint" cannot be captured by external inspection. Many "imprinted" processes are stored with little conscious awareness of what was learned, so that an individual cannot tell another person about the contents of such an imprint. One example of low awareness "imprinting" involves viewing pictures. A very large database of images could be used, from which certain image(s) could be selected for viewing. If an individual were to view a previously shown image, grouped with another image that had not been previously shown, the individual could select the previously seen image with high confidence, even long after the initial training.

One use of imprinting is found in the work of Dhamija and Perrig (2000), who have the user select a small group, their portfolio, of images from a larger set of images. Recognition of these images certifies the user. The taught protocols emphasize making the user aware of the selected images, and using the same images repeatedly for identification. The motivation was to achieve more natural human factors, "pictures replacing passwords," at a modest security level. However, repetitive use of pictures could easily lead to similar problems as for regular passwords, namely that an eavesdropper could steal such a "picture password".

A scheme recently described by researchers at Microsoft (Microsoft) uses cued recognition of artificially generated Rorschach patterns to generate passwords which would be too long to remember and impossible to guess. The user is shown a set of pictures and asked to assign a word to each, keeping it secret. Letters selected from these words become the password for subsequent certification. The pictures provide cues to recall the chosen words, and thus the passwords. Again it appears that the evaluation that is done is of the password cued by the pictures, and does not involve a probabilistic assessment of error. Furthermore, it is still possible to steal the images, which are used repetitively, and/or to otherwise attack the password itself.

SUMMARY OF THE INVENTION

The background art does not teach or suggest the use of imprinting as part of a security function that is performed with the human user and that is required for a security and/or identification mechanism. The background art also does not teach or suggest the use of imprinting as a cryptographic primitive.

The present invention overcomes these disadvantages of the background art by providing a wide range of human memory *imprinting* phenomena as potential cryptographic primitives. These "imprinting" phenomena are characterized by vast capacity for complex experiences, which can be recognized without apparent effort and yet cannot be recalled directly. Thus they become natural "one-way functions" suitable for use in near zero-knowledge protocols, which minimize the amount of secret information exposed to prying eyes while certifying an individual's identity. It should be noted that this use of imprinting is not intended as a simple "picture password", but rather uses imprinting as part of a security function that is performed with the human user.

The term "imprinting" is used herein to describe the range of memory phenomena in which the information stored greatly exceeds the amount which is easily recalled by a person, and in which the information is stored with little or no perceived effort. In the psychology literature, these phenomena are broken down into two, three or more categories, and the boundaries are not always clear. For example, there are studies of implicit learning, procedural learning, or "priming." The process of the present invention is an example considered to be implicit learning. Learning to ride a bicycle is the classic example of procedural learning, although there are things other than motor skills which are learned by the procedure of doing them multiple times. Priming usually describes phenomena of which the individual is completely unaware. Imprinting may optionally include one or more of these categories, but is more preferably directed toward implicit learning.

Cryptographic primitives may be considered to function as follows. For the present invention, human memory is considered to loosely resemble a one-way function. One certainly cannot run it backwards to extract what has been stored for purposes of telling another person what that is. A one way function is a transformation which is easy to carry out but cannot be reversed without expending an unrealistically large computational effort. Thus even if an adversary has the encrypted message and the key used to encrypt, and knows the function used to encrypt, it is not possible to determine the original message.

The present invention also preferably uses a plurality of pictures or other items capable of being sensed for imprinting. More preferably, as described in greater detail below, each picture (or other imprint) is used only once, as for the one-time pad. The one-time pad is a type of encryption in which an encryption scheme depends on a sequence of random numbers, each number used to encrypt one symbol and then discarded, never to be used again. No method of guessing frequently-used patterns in the message may be used when the code is being discarded as fast as it is used.

The present invention also preferably structures the protocol to expose the fewest possible portions of the imprint in each session, using the same ideas as near-zero knowledge exploits, namely the probabilistic assessment of the likelihood that this is not an authorized user, but an imposter, stopping when this probability drops below some prearranged threshold. Zero-knowledge or near-zero knowledge protocols are not usually used to encrypt whole messages, but to certify some fact without actually revealing its details. For example, a zero-knowledge proof may be conducted between two parties in a series of rounds. Party A wishes to prove some fact to the satisfaction of Party B by answering the questions of Party B, which Party A could only answer if the fact is true. Party B accumulates enough evidence about the truth of the fact of Party A in several rounds to convince Party B that Party A is telling the truth. Party B does not obtain the details of the secret fact in this way, and neither does any eavesdropper.

These functions or characteristics of the human brain include the following. Human memory has the capacity to quickly learn vast amounts of information (pictures and strings). This capacity allows the use of cryptographic zero-knowledge-like authentication protocols, which minimize the exposure of the shared information upon which the certifying transaction is based. Such protocols rely on the probabilistic evaluation of acceptance error (the likelihood of false identification), and are safe from eavesdroppers, since only a few bits of information are securely transmitted, and those bits are used only once.

Also, the stored information is hard for people to recall but easy to reveal by less direct means long after the initial presentation. One example is recognition: users are asked to recognize an example of the material as one to which they have previously been exposed, rather than to recall an object (a "shared secret") unassisted from memory. The authentication protocols of the present invention preferably access human memory without the need for recall, which makes the protocols more pleasant to use and safer from imposters: the knowledge required for authentication cannot be passed from one person to another.

The validation of imprinted certificates is inherently a probabilistic process, since it involves human performance. The present invention also includes methods for certifying a user, analyzing in each case the protocol required to reduce the chance of imposture by guessing or eavesdropping to some desired small probability. The present invention may optionally include tasks related to any cognitive-sensory function, including but not limited to, verbal tasks, visual tasks, olfactory (smell-related) tasks, audio tasks, taste tasks or touch-related tasks. Optionally users may be allowed to select a particular sensory protocol, for example depending on whether their memory is more suited for sounds, words, images etc. All the phenomena are described in the relevant literature of perception and cognitive psychology, where the basic effects are not controversial (although the underlying mechanisms may be (6)). However, the authentication protocols are new and inventive, since certification is a new application that was not discussed in the art.

According to the present invention, there is provided a method for providing a security function with a user, comprising: imprinting the user with at least one cryptographic primitive determined from a sensory mechanism; and at least one of authorizing, identifying or authenticating the user according to an ability to recall the at least one cryptographic primitive.

Preferably, the imprinting comprises implicit learning by the user. More preferably, the at least one cryptographic primitive is used to encrypt a message according to a one-way function. Also more preferably, a one-time pad comprises the at least one cryptographic primitive. Also more preferably, a near-zero knowledge function comprises the at least one cryptographic primitive.

More preferably, the sensory mechanism comprises vision, such that the at least one cryptographic primitive comprises recognizing an image. Most preferably, the recognizing the image comprises: training the user on a plurality of trained images; and testing the user on a combination of a trained image with at least one distractor image. Also most preferably, the at least one distractor image comprises a plurality of distractor images.

Preferably, the testing comprises: selecting a plurality of different trained images by the user in sequence, the sequence providing the cryptographic primitive for determining the at least one of authorizing, identifying or authenticating the user.

According to another embodiment of the present invention, there is provided a method for authenticating, authorizing or identifying a user, comprising: training the user with information through a sensory mechanism; and determining accurate recall of the information to authenticate, authorize or identify the user.

According to yet another embodiment of the present invention, there is provided a method for a one-way function for authenticating, authorizing or identifying a user, comprising: imprinting the user with a cryptographic primitive; and testing the imprinting with at least a similar or identical cryptographic primitive to authenticate, authorize or identify the user.

Preferably, the cryptographic primitive is derived from input according to a sensory mechanism. More preferably, the input comprises at least one image and the sensory mechanism is visual.

Also more preferably, the input comprises at least one pseudoword and the sensory mechanism is verbal.

Preferably, the sensory mechanism is selected from the group consisting of tactile, olfactory, audible and taste.

Also preferably, the testing comprises determining whether the user is capable of discriminating between an imprinted cryptographic primitive and a non-imprinted cryptographic primitive.

Preferably, authorizing, identifying or authenticating comprises measuring user knowledge of one or more facts of said cryptographic primitive via a query presented to said user. The use of these one or more facts and user knowledge thereof means that the answers to the query are not present in front of the user on the screen and therefore an eavesdropper cannot understand the answers that he sees. Typically, the imprinting said cryptographic primitive comprises learning an image, and one of said facts is that a given image is a learnt image and a second of said facts is associated with a position of said learnt image in a grid. Alternatively, a second of said facts is associated with learnt changes in said image.

The method may involve repeating said measuring for different cryptographic primitives for a predetermined number of times, or until a given level of certainty has been reached.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention is herein described, by way of example only, with reference to the accompanying drawings. With specific reference now to the drawings in detail, it is stressed that the particulars shown are by way of example and for purposes of illustrative discussion of the preferred embodiments of the present invention only, and are presented in the cause of providing what is believed to be the most useful and readily understood description of the principles and conceptual aspects of the invention. In this regard, no attempt is made to show structural details of the invention in more detail than is necessary for a fundamental understanding of the invention, the description taken with the drawings making apparent to those skilled in the art how the several forms of the invention may be embodied in practice.

In the drawings:

FIG. 1 shows an exemplary dual perception image;

FIG. 2 shows an exemplary closure image;

FIG. 3 shows a flowchart for an exemplary method according to the present invention;

FIG. 4 shows a flowchart of an illustrative security implementation of the method according to the present invention;

FIGS. 5A and 5B show graphs which compare the behavior of three model users according to different simulations of the method according to the present invention;

FIG. 6 shows a graph with results from actual users trained with the method according to the present invention;

FIG. 7 shows a graph with results from actual users concerning recognition accuracy for the method according to the present invention as implemented with pseudowords;

FIG. 8 shows a finite state machine (FSM) which generates a "grammar" of strings;

FIG. 9 is a simplified diagram illustrating a flow chart for constructing a multiple reuse embodiment of an imprinted certificate according to a preferred embodiment of the present invention;

FIG. 10 is a simplified diagram showing a checkerboard matrix for use with the flow chart of FIG. 9;

FIG. 11 is a simplified diagram showing a matrix for use with a three-hidden groupings embodiment of the present invention;

FIG. 12 is a simplified diagram showing a four by five image matrix as used in an implementation of the present invention;

FIG. 13a illustrates an image and its pair in accordance with an embodiment of the learning or imprinting part of the process of the present invention;

FIG. 13b illustrates a greyed image and three greyed variations thereof in one of the implementations of the learning procedure of the present invention;

FIG. 14 is a graph of success rate against number of days for three subjects undergoing learning procedures according to the presently described embodiments; and

FIG. 15 is a graph of imposter acceptance rate against days of training for different kinds of query according to preferred embodiments of the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is of a system and method for using imprinting as part of a security function that involves a user, for a security and/or identification mechanism. Imprinting is preferably used for cryptographic primitives, for determining a one-way function that operates at least partially according to a characteristic and/or function of the human brain.

The present invention has a number of advantages over the background art. For example, recognized images from a small set of pictures have high user awareness, yet are difficult to describe to another person. As a result it is difficult for another person to impersonate the user by correctly identifying images. Nonetheless, eavesdropping will succeed if the same image (or few images) is always used as a certificate or password. The effort to remember an image is less than the effort to remember passwords, and the verification protocol is more complex, but not much so. The present invention further increases the security of using images for the security function by not using a particular image as a password. Instead, preferably information is learned automatically in a process known as "priming", for which examples are described below. The information is stored in a procedural memory of the user. The user is unaware of the ability to identify the imprinted information, and cannot describe it to another. Therefore the likelihood of successful impersonation is extremely low, and eavesdropping is not a danger. No effort is required to retain the "primed" ability, but the measurement of these effects may be subtle, and require the most complex protocols.

The present invention relies upon particular cognitive functions or characteristics of the present invention, such as the distinctions between explicit and implicit memory, and recall (free and cued) vs. recognition. These distinctions are not always fully understood, but enough is known that can be used for the design of many types of certificates (information for security functions), with different properties that may be appropriate for different applications.

Procedural memory covers a range of distinct memory abilities which rely on information that cannot be verbalized. During the acquisition of procedural memory, what is learned cannot be described. Typical examples are motor skills (such as riding a bicycle) and perceptual learning, but the acquisition of grammar is usually also considered to be a process of procedural memory. A related concept is that of implicit memory, where memory of past experience is retained without specific recognition of an invoked memory of past events. Manifestations of implicit memory are typically called priming, and they can come in different modalities, including image priming and string repetition priming.

A common characteristic of both procedural memory and implicit memory is that recognition of an invoked memory is relatively effortless, requiring no or little conscious effort. Such memories can be retained effortlessly for a very long time, and be invoked with little preparation when the need arises. Moreover, since these memories cannot be verbalized or easily described, they cannot be easily revealed to other people, whether willingly or unwillingly. All of these characteristics are very useful properties for a verification certificate, which provides information for a security function. On the negative side, it may take time to acquire procedural skills or become primed by implicit memories, and the verification may also require a relatively laborious protocol.

These characteristics of human memory are heavily used for advertising. It is commonly assumed that mere exposure to names and labels increases people's willingness to buy "familiar" products in the future. In a more controlled environment, it has been shown that objects previously exposed in advertisements are judged as 'more attractive' by people, even if the advertisements were not carefully watched during the time of exposure.

As an example for the use of procedural memory for "imprinting", the paradigm of artificial grammar learning can be considered, initiated by Reber (Reber 1967). In this paradigm people are asked to memorize a list of words generated by a finite-state automaton. After the initial training period, the subjects are told that the words were generated by a certain grammar, and are asked to recognize new words as "grammatical" or not. Subjects perform better statistically than chance "guessing" on this task, and better than control subjects who have not observed the "grammatical" strings before. When asked to describe how they perform such discrimination and identification, subjects are not able to describe what they have learned and what rules they are using to do the task. For this reason this task is considered by some to involve implicit memory.

As an example of priming, the paradigm of ambiguous figures can be used. For ambiguous pictures, two percepts are possible but not simultaneously, such as the famous picture shown in Figure 1, which may perceived as showing a young girl or an old woman. Previous exposure to the stimulus which favored one of the possible perceptions of the picture "primes" people to prefer that same perception in later exposures (Long and Olszweski, 1999).

Another possible example involves 'closure pictures', somewhat similar to the famous “Dalmatian dogs" picture shown in Figure 2. With previous exposure to such pictures shown with increasing level of details, recognition in a second exposure can be done from a "sparse" picture. In such a picture, only a few details (for example, only fraction of the edges) are presented. Untrained subjects cannot recognize such pictures. Thus subjects are primed to recognize pictures of objects they had seen before from very few details, but are not able to do the same with new objects.

Additional examples for identity/repetition priming include but are not limited to, a fragment completion task: Tulving et al (1982) asked participants to learn long, infrequently used words e.g. TOBOGGAN. Either 1 hour or 1 week later participants are asked to fill in spaces on a page in a fragment completion task (_O_O_GA_). Participants exhibited repetition priming, such that performance was better on words seen previously. This is a very long-lasting effect, and lasts up to months.

Another example is picture naming: (Cave 1997) reported faster naming of pictures repeated from a prior exposure than new pictures; this effect lasts up to a year or more. Musen & Treisman (1988) showed that a single exposure of a novel, nonverbal stimulus supports long-lasting perceptual priming, while recognition memory rapidly deteriorates.

Perceptual learning is another example, as many low level visual skills (such as texture discrimination or even contrast detection) can be improved with practice; improvement tends to be very specific, which is useful for the purpose of priming a certificate, and can last a long time (e.g., 2 years in texture learning, see (Karni and Sagi, 1993)).

Explicit memories also come with different degrees of awareness. Some memories are easy to recall freely (such as an individual's name and address), but most explicit memories require some assistance for recall. Pictures, for example, are relatively difficult to recall but easy to recognize. Regardless of the memory modality, it is almost always easier for people to recognize items in memory (for example, indicate whether a selected item is familiar or not) rather than freely recall items from memory. In between recall and recognition, there is the "cued recall" paradigm, where groups of items (such as words) are associated with each other, so that when encountered with one item, a person can easily retrieve the second or more items. This characteristic of human memory is commonly used in the teaching of languages, and possibly other skills.

One important characteristic for the design of certificates is that explicit memories are relatively easy to acquire. As long as the recognition or cued recall paradigms are used, it is still often the case that the memory traces are hard to verbalize or otherwise be given away to other people. Also, if the modality, selected from a sensory function, is suitable for such memories, such as images being used for the visual sensory function, only a little effort is required to maintain a very large store of items; these items can then be used for verification on a one-time basis, and the verification protocol can therefore be safe from eavesdroppers.

One non-limiting example of a low-awareness recognition process is the use of a large database of images for establishing visual imprinting with a user. It has been shown that people can remember a very large number of pictures following a single short exposure to each picture. For the present invention, optionally multiple short exposures are used to consolidate the memories. Visual memory (memory for pictures) is potentially very long term, lasting up to years (Sheppard, 1967) and the capacity appears limitless (Standing et al, 1970). As an example of cued-recall in another modality, the paired associate paradigm may optionally be used, in which users are asked to provide a matched group word/figure to a given cue word/figure.

Another non-limiting example involves an effect known as ‘change blindness’; recently a number of studies have shown that under certain circumstances, very large changes can be made in a picture without observers noticing them. In these experiments changes are arranged to occur simultaneously with some kind of extraneous, brief disruption in visual continuity, such as the large retinal disturbance produced by an eye saccade, a shift of the picture, or a brief flicker, e.g. Rensink et al (1997). However, once a subject becomes aware of the change, it is rapidly perceived in subsequent viewings.

Preferably, training for the present invention is performed in at least one session but more preferably in more than one session; if a plurality of sessions is used, they are preferably performed on successive days. Depending upon the complexity of the material on which the subject is to be trained, the training session may range in length from a few minutes to a few hours, but preferably is of relatively short duration (such as up to about one half hour for example). A sufficient large set of objects on which imprinting is to occur, such as pictures for example, is preferably used; for example, for pictures, a set of pictures ranging in size from about 10 to about 500 pictures was used; a medium size set (for example around 100 pictures) was found to be preferable. The database of pictures or other objects may optionally be of any size, but is preferably from hundreds to thousands or even millions of objects.

The training session is optionally and preferably ended with a short practice test session. More preferably, a test session is performed shortly after completing the training session but with a break of from a few minutes to a few days. Refresher training may optionally be performed as needed, depending upon the number of objects in the original training session and the rate of use.

For an actual test session, preferably a plurality of distractors is used, since it was found to decrease the chance of someone from guessing the correct object (picture etc) without appreciably decreasing the chance of trained individual to identify the correct object (data not shown).

The principles and operation of the present invention may be better understood with reference to the drawings and accompanying descriptions.

Before explaining at least one embodiment of the invention in detail, it is to be understood that the invention is not limited in its application to the details of construction and the arrangement of the components set forth in the following description or illustrated in the drawings described in the Examples section. The invention is capable of other embodiments or of being practiced or carried out in various ways. Also, it is to be understood that the phraseology and terminology employed herein is for the purpose of description and should not be regarded as limiting.

EXAMPLE 1

VISUAL PERCEPTION AS THE SENSORY MECHANISM

This Example describes an illustrative method for certificates based on visual recognition, by using a very large database of images for training, followed by recognition of at least one image. Recognition of images forms the cryptographic primitive, with visual perception and recognition as the sensory mechanism for the method of the present invention. Preferably recognition is performed in the form of discrimination, such that the user is able to select a correct image from a plurality of images. The correct image may optionally have been shown previously, or alternatively may be similar to a previously displayed image during the training process.

The exemplary method of the present invention, shown in Figure 3, preferably starts with a training session in stage 1, during which the user is shown a relatively large set of images, preferably randomly selected from a very large database, preferably at least about 100,000 images, although there is no upper limit. Image databases with 1 million or more pictures exist already, and larger ones are coming into use as digital photography becomes more prevalent. An important, limiting practical issue is therefore the ability to select groups of pictures for this use that are easily remembered, have a common central figure or story, and are not so similar as to be confusing, rather than the database size. This process may even be performed manually, as was done for this Example, with some simple tools to record the choices of images. It was found to be possible to construct 500-1000 groups from a much larger database, and then select randomly the one picture in each group which a particular user would be trained on, reserving the rest of each group for use as distractors.

In stage 2, after the training process has finished, the memory of the images may optionally be used for authentication. During authentication, the user is shown a small set of preferably randomly selected images (preferably from about 2 to about 9 images) side by side, only one of which was present in the original training set. In stage 3, the user identifies the image shown during the training session, and/or the most similar image to one displayed during the training session. This stage may optionally and preferably be repeated more than once, to defeat random guessing. To defeat eavesdropping, each image in the training set is optionally and more preferably used only once for certification (security and/or identification and/or authentication) purposes. Thus retraining is preferably performed when the trained set of images is exhausted.

To analyze the effectiveness of picture recognition as a certificate, for performing the above security function, the behavior of an imposter who has not been trained on the same specific images is considered. Let n denote the number of images shown side by side in each trial. The imposter would guess correctly 1/n of the time. The performance of the user might also not be perfect, but can be distinguished from guessing on a statistical basis after a few presentations. A certification application can optionally operate by presenting images for recognition and stopping as soon as the chance that guessing would have produced the observed number of correct recognitions is reduced below a preset threshold.

As described in greater detail below with regard to Figures 5A and 5B, the number of trials which are required to certify that a user who correctly recognizes the trained information (picture, pattern, pseudoword etc) a certain fraction of the time is not a random guesser was calculated. There are two other parameters to control -- the number of "distractors" which are presented and the tolerable acceptance error. In Figure 5A, the tolerable acceptance error is 0.01 or one chance in 100 that the entries were made by a guesser. In Figure 5B, the tolerable acceptance error is 0.001 or one chance in 1000 that the entries were made by a guesser. The number of distractors considered is 1 for one set of lines and 6 for the other, that is, two patterns were shown in each presentation for the first set of data, and seven patterns were presented in each trial for the second set. Finally, because the user makes errors at random, the result is a distribution of success rates, so a cumulative distribution is shown. The vertical axis is the probability that a user with a particular accuracy is accepted in N or less trials.

Figure 4 shows a flowchart of an exemplary security implementation of the method according to the present invention. As an example of the use of imprinted behavior as a certificate of identity, assume that a portable computer is to be protected from unauthorized use, for example to block an unauthorized individual from turning the computer on, logging in as the authentic (permitted) user, and accessing stored information therein. The application that controls security preferably uses the certification method according to the present invention.

As shown with regard to Figure 4, in stage 1, a database of images (pictures) is preferably provided, more preferably at least about 100,000 such images. Optionally and preferably, the images are stored on the portable computer, for example on the hard disk of the computer. The pictures are preferably organized in groups, with a common theme, preferably also having a common focal point or narrative, such as two or more wild animals, two or more landscape scenes, two or more city scenes, etc.

In stage 2, the user is trained with the images from the database to be identified in the future. Optionally, the training program is operated by the portable computer itself, using the database of images; alternatively, the training may optionally be performed in some other way, outside of the operation of the portable computer. The training process preferably includes presenting a large number of pictures from the database, more preferably selected at random, for a short period of time, optionally 5 seconds or so apiece. Only one image of each group is preferably used for training.

In stage 3, the displayed images are marked in memory or otherwise noted, by the application as operated by the portable computer and/or by another external application.

In stage 4, the user is to be authenticated as having authorized access to the computer. Preferably, the user is shown groups of pictures from the groups that are stored together in the computer database, one of which has been shown to the user before.

In stage 5, the user selects an image that has been shown before, preferably one image of a group of images, only one of which was displayed during the training session. In a group of k images, the chance of an imposter (individual who is guessing) being correct is 1/k.. Even for k = 2 after 6 trials the imposter's chances of being correct every time are less than 1 in 50, but for k = 7, the imposter is expected to guess correctly in four successive trials less than one time in 2000. Thus, if the process is repeated, then the chance of guessing correctly is reduced significantly. Even if the user makes an error occasionally, say one time in 10 trials (the literature and the inventors' experiments suggest that a higher degree of accuracy can be maintained), 10 trials would be sufficient to reduce the probability that the performance could be produced by guessing to between 1 in 100 and 1 in 1000, as described herein, even if only a pair of images is used. The user may optionally set the desired level of security, such that the authentication program would test the user only until the user has performed the authentication process to the desired degree of certainty, according to the formula with which Figure 5 (5A and 5B) has been calculated (see below). This saves time, and exposes the fewest pictures to possible "eavesdroppers".

The present invention also preferably includes a method to protect the authentication/authorization application from viewing the image(s) by looking over a user's shoulder or otherwise gaining unauthorized visual access to the image(s) during the authentication process. One option is use an image from the database only once. Alternatively, the user may be asked if the process occurred unobserved, such that the images could optionally be used again.

Optionally and more preferably, when insufficient images remain for the authentication process, the user is trained with more images from the database, and/or another database is provided, after which the training process is performed again, as shown with regard to stage 6.

A similar method may optionally be used when the database is on a server in a central, secure location, and the person desiring to be verified communicates with the system over a communications channel. The communication may be recorded, so images used for remote certification are preferably not reused at least for this purpose. Retraining to add extra images when the supply is low is preferably performed in a more secure location, such as on the user's personal computer for example.

Figures 5A and 5B both compare the behavior of three model users. In Figure 5A, the cumulative distribution is shown of the number of trials required to reduce to 0.01 the chance that an imposter, guessing, could impersonate a valid user. In Figure 5B, the chance of guessing (tolerable acceptance error) is reduced to 0.001. The dashed lines represent a protocol with two choices, the solid lines a choice between 7 alternatives. The three model users have, on average, 95% correct, 90% correct and 80% correct performances, such that they make errors at a constant rate of 5%, 10% or 20% of the trials, respectively. Two scenarios are considered: n = 2 and n = 7. The model stops presenting pictures for recognition when the chance that an imposter, guessing randomly, will do as well as the user has been reduced to 1 in 100. This is usually accomplished within three trials in the seven choice scenario. Only the least accurate user will ever require more than 6 trials to reach this level of certainty. If the threshold of certification is set at 0.001 (Figure 5 B), this protocol would require 5-7 trials under the same range of assumptions about user performance. Over this range of user performance, the 7-choice protocol requires 3-6 trials to certify identity at the 1% level, while the 2-choice protocol requires 7-11 or more.

Figures 5A and 5B show cumulative distributions. Each curve gives the probability that a user making errors at a specified rate will nonetheless reach the desired certification threshold at or before the number of trials indicated on the x-axis. Introducing more distractors makes it harder for the opponent to fool this system, since with one distractor, certification is not always obtained with high accuracy, while with 6 distractors, the model shows that high accuracy can be obtained always or at least with a very high frequency. If there are sufficient distractors (the group of lines on the left side of Figures 5A and 5B), the use of two more trials provides ten times more power in rejecting an imposter.

In a binary forced choice protocol (the dashed lines of Figures 5A and 5B) there is a greater premium on user accuracy. The user who makes 20% errors may require 20 or more trials before the system will certify this user at the 1% level, an unreasonable amount of effort. Although one might suspect that presenting more choices might cause users to make more errors, actual experimental results (not shown) found that the decrease in accuracy is slight or absent, so that increasing choices seems always to be a good design decision.

As shown with regard to Figure 6, actual experiments confirm that subjects, trained on 100-500 pictures in a training session lasting from a few minutes to half an hour, were often able to recognize previously seen pictures with better than 80% accuracy for at least a month and often much longer. The process was first studied with three subjects who were presented with a previously seen image and one not seen which were similar in most of their elements (e.g., two pictures of giraffes, one with two and the other with three giraffes). This proved more confusing than helpful to the subjects. Their performance, initially high, began to deteriorate to 70-80% after a month or two. When pictures were selected more randomly, by selecting pictures with a clear central subject or action, performance improved to that shown in Figure 6. Using the same methodology for picture selection, we are now presenting subjects with 6-9 choices of picture. Preliminary results suggest that recognition percentages are as good as or better than were achieved with binary forced choice presentation.

Figure 6 shows recognition accuracy achieved by three subjects, each trained on a fixed set of 100-500 pictures and then asked to select the previously-seen picture from a group of pictures at various later times. No trained picture was presented more than once in the testing. The two data files labeled "69pictures" are subjects shown pictures in groups of 6 to 9.

According to another optional embodiment of the present invention, rather than using every group of pictures (including one picture from the training session and the rest as distractors) only once, a variant method is optionally performed in which the pictures (optionally including both images on which the user was trained and also distractors) may optionally be used multiple times. Experiments with a number of human subjects have shown that it is possible to reuse these patterns, as they are more familiar when a subject sees them a second or third time (or more), while the distractors do not appear to also become familiar when reused (data not shown). These experimental results (data not shown) also indicate that recognition accuracy increases when groups of images (one trained picture plus related distractors) are reused, and that repeated exposure to the distractors does not confuse the user.

Therefore, reuse of the patterns is possible, but carries some exposure to eavesdropping. Preferably, the method includes safeguards against an eavesdropper being capable of understanding, guessing or otherwise obtaining the underlying pattern or other information concerning the encoding method.

EXAMPLE 2

VERBAL PERCEPTION AS THE SENSORY MECHANISM

This Example relates to the recognition of pseudowords, in which recognition of the pseudowords forms the cryptographic primitive, and verbal perception and recognition is the example of the sensory mechanism. A recognition protocol can also optionally be designed with strings of letters, when it is not possible or not desirable to use pictures because of the additional memory and storage required, or because an adequate display facility is not available. Instead of pictures, this implementation of the present invention uses pseudowords, generated by taking a list of over a thousand common English words obtained from Wilson (10), and modifying them in one letter position using the program provided at (11). A native English speaker then selected pseudowords which are pronounceable, and do not exist as valid words. In this construction, the method followed a protocol similar to the one used in (12). Of course, the method could optionally be extended to form pseudowords in any language by a similar method.

As with pictures, during training subjects are familiarized with a random set of pesudowords. During verification subjects are presented with a plurality of pseudowords, preferably only one of which has been shown to them before, and are asked to identify the previously trained pseudoword. Pseudowords differ from pictures in that the native language of the user is expected to have an effect on the user's ability to recognize pseudowords based on one language. As an advantage, the pseudowords are expected to require less storage or transmission time than pictures. Recognition rates obtained with pseudowords in experiments, as shown with regard to Figure 7, are comparable to but not quite as good as the accuracy seen in recognizing pictures. The results are shown for two subjects, with the pseudowords presented in groups, including one pseudoword shown during the training session and one not shown.

The picture recognition protocol has some advantages over the method of the present invention with pseudowords; for example, the picture recognition protocol is easier to use; it is more or less universal across cultures; and people demonstrate rather good long term retention of the pictures. Pseudowords are harder to train and somewhat less reliable, but they can be used when pictures are not an option. In order to maintain the safety of the protocol from eavesdroppers, re-training with a new set of pseudowords is preferably performed when all the training examples are used, as for the image recognition protocol.

EXAMPLE 3

SKILL ACQUISTION AS THE SENSORY MECHANISM

This Example relates to certificates or training toward a cryptographic primitive that is based on skill acquisition, in which performance of the acquired skill represents the cryptographic primitive.

Skill acquisition may optionally be performed as based on the AGL (Artificial Grammar Learning) task first introduced to the literature of cognitive psychology by Reber (1967). In his experiments, subjects first learn sets of approximately 20 strings of three to eight characters. Although these letter strings might appear random, they have, in fact, been generated by a Finite State Machine (FSM) such as the one used by Reber, shown in Figure 8. Figure 8 shows a finite state machine (FSM) which generates a "grammar" of strings.

To produce a string with an FSM of Figure 8, one begins at the "start" arrow at the left, and traces around the diagram in the direction of the arrows until one reaches "end" on the right. Each transition from one state (circle) to another generates a letter which is added to the end of the string generated thus far. One can either construct all possible strings and sample from that set at random, or associate probabilities with the possible directions one can take at each node of the FSM, and in that way associate a probability of being generated with each possible string. The first method was used in the initial experiments by the inventors, although both are potentially useful for the purposes of this invention.

For example, the strings possible with this particular FSM include:

TTS
VVS
TPTS
VXVS
TTXVS
VVPS

Reber's (1967) main finding was that his subjects could memorize "grammatical strings" which were systematically generated (by the FSMs) more readily than they could learn truly random strings. Still, the subjects were typically reported as unable to articulate the patterns they had learned. Even after being told explicitly that the strings they had learned were governed by "a complex set of rules" they were unable to give anything but the vaguest characterization of the strings' structure. Nonetheless, when given a forced-choice task with strings they had not previously seen, they were able to correctly distinguish strings that had been generated by the same FSM from random strings at a rate of nearly 80%.

The AGL task may therefore optionally be used as an "imprinted certificate" for the purpose of the present invention, in order to train subjects. Optionally and preferably, longer strings of 3-10 characters are used, and a larger FSM with 8-10 internal nodes to generate them.

For each user requiring a certificate, an FSM is created at random.

The strings are preferably presented to the subject for identification, grouped or shown with at least one other string that is generated by an FSM which is similar in structure but has one or a plurality of letters in the wrong positions, optionally and more preferably at interior locations in the string. The experiments performed by the inventors showed a subject performance of better than 90% accuracy, even with this difficult choice, for short periods of time, and continued performance at better than 60% over several days was demonstrated. This degree of accuracy is sufficient to separate the real individual from an imposter, but requires more trials than the picture recognition test. However, because of the greater human error rate, the likelihood of an eavesdropping computer understanding how the strings are generated is enormously less likely. The convenience of the test can be increased by making the comparison string(s) (the wrong choice in each group) random or more nearly so, at some decrease in security.

Reference is now made to Fig. 9, which is a simplified flow chart illustrating a further preferred embodiment of a secure authentication protocol intended for multiple reuse of the imprinted memories. In respect of Fig. 9 and the succeeding figures is described a secure authentication protocol which relies on the same skill of picture recognition as in the preceding embodiments, a skill which people find relatively easy. The human and the computer share a secret, which is a set of 60-100 pictures. Authentication is done via a challenge-response protocol: the computer poses a sequence of challenges to the user, which can only be answered correctly by someone who knows the shared secret. Once the probability of random guessing goes below a fixed threshold, the computer authenticates the user. We report user studies showing that the protocoi is reasibie ior humans to use, with high reliably and for a long period of time. We also describe probabilistic attacks on the protocol, which demonstrate the protocol's computational merits and limitations.

As discussed above, the intense interaction between humans and computers in recent time has made traditional areas of cognitive psychology and perception highly relevant for the emergence of new technologies. But at least one area in computer science has managed to shy away from any such influence - the area of computer security and authentication. We are still using highly non-user-friendly passwords, and most security protocols are developed intended for use by computers (or humans assisted by computers).

REUSE EXAMPLES

In the following embodiments we continue to address the problem of user authentication. We focus on authentication done over insecure networks from potentially compromised computers, such as in internet cafes. In such cases there is a high risk that an eavesdropping adversary (EVE) records the communication between the user and the main computer. It is therefore necessary to develop secure authentication protocols, where overhearing one or more successful authentication sessions will not let EVE pose as the legitimate user at a later time.

Clearly our everyday password is not secure in the sense we require - by merely recording the input of the user to the intermediate computer, Eve can discover the user's password after a single successful authentication session. Biometric identification (based on such physiological traits as fingerprints and iris shape) is indeed more secure against theft or forgetting, but it is just as easy for Eve to obtain this key as it is for her to obtain a password.

The only existing secure solution is for the user to carry a computational aid, such as an OTP card that generates one time passwords, a laptop armed with secure authentication protocols, or a simple transparency. But this approach has its drawbacks, for example users cannot get authenticated without the device, which can be stolen, lost, or made unusable (e.g., when its battery runs out).

The present embodiments intend to develop a user authentication scheme that is secure against eavesdropping adversaries, and yet can be used reliably by most humans without tne neea or any external computational aid. Not much has been said about this problem. Recent systems have been developed which use easy to remember passwords, such as a small number of faces as in the commercial Passface TM System, abstract art pictures, or memorized motor sequences. These schemes use passwords that are indeed easier to remember, but they are not safer than regular passwords against eavesdropping adversaries. A few recent cryptography papers tried to address this issue, but their proposed protocols are either not secure for any sufficient length of time, or impractical in that most humans cannot use them.

In our previous work, we have proposed an approach which is motivated by insights from perception and cognitive psychology. Basically, we studied a variety of memory modalities including pictures and pseudo-words, taking advantage of the vast capacity of human memory to design protocols that use each memory item only once, and are thus perfectly safe against eavesdropping. These protocols have the added value (or drawback) that they cannot be ``loaned" by legitimate users to other people. The main drawback of these protocols is that with extensive use they require frequent retraining of users.

The present embodiments are intended to provide an authentication protocol that is relatively easy for most people to use without any computational aid, that is safe against eavesdropping adversaries even after the completion of a large number of successful authentication sessions, and that does not require frequent retraining.

In the preceding embodiments, the preferred use of the idea has been to create a sort of "one-time" certificate that exploits our ability to retain and recognize a large amount of natural information, without having a very detailed knowledge of what we have learned. Each piece of information is used only once, so that an eavesdropper learns nothing of use if he has access to each of our certification sessions. The drawback to this is that eventually it will be necessary to train the user on fresh patterns as the initial set is used up.

As indicated above, if the user's answers in a recognition interaction are disguised in certain ways, the danger of an eavesdropper learning which patterns are parts of the certificate can be eliminated. Here we describe several ways to do this. In each version described here we train users on a group of images that is uniquely chosen for that user, stage 100. Subsequently, in stage 102, we place, typically, two or three trained images in a grid as part of a group of 16 or 18 pictures, the other pictures serving as distractors. In a stage 104 the trained images are associated with one or more sub-groups in one of several ways that will be discussed in greater detail hereinbelow, and the grid positions are associated with indices, preferably separate indices for the separate sub-groups, as will be discussed hereinbelow. The grid with the images is shown to the user in a stage 106. In the previous embodiments we analyzed the results of presenting one trained image and n-1 distractors, such that an imposter's chance of correctly guessing each image is 1/n. The imposter's chances of correctly guessing 2 out of n images is 2/(n*(n-1)). Presenting two images at a time with twice as many distractors is thus inherently about twice as strong as presenting each image separately with its share of the distractors. The user recognizes the familiar images and selects the correct indices. For two images there should be two indices to add together. The result is entered by the user in stage 108 and the entire process is repeated either a given number of times or until a certain probability of exclusion of imposters is reached.

Reference is now made to Fig. 10, which is a simplified diagram illustrating an image grid 120 according to one preferred embodiment of the present invention. Suppose we present 18 pictures at a time, including three previously trained, arranging them on the squares 122 of a checkerboard with the black and white colors of the checkerboard indicated. Associated with the array of pictures are two grids of numbers 124 and 126, arranged in the same pattern as the pictures. Each grid contains the numbers 1 through 18, but in a random permuted order. The user may be asked to indicate whether the majority of the "familiar" pictures lie on the black or on the white squares of the checkerboard. If there are more black than white familiar pictures, use table 124, otherwise pick table 126. Now add up the three numbers found in the same position in the table chosen as the pictures previously trained appear in the picture layout, and report as the answer the final digit (the sum, modulo 10).

Now the chance of an imposter guessing the final result is 1/10, rather than the smaller chance that wouild have resulted if the user reported all of his observations. There is little chance that the observer will learn from this obscured answer just which pictures were trained on, since new random rearrangements of the numbers 1-18 will be used for each interaction. With perfect recognition by the user, three trials will give 10^-3 chance that an imposter will be accepted.

A variant method is now described with respect to Fig. 11, and involves learning additional information associated with each picture. In our current trials with this scheme, we train subjects on 36 images, in three groups of 12, and ask them to remember which group a particular image was a member of. Thus each image intrinsically has a label, 1, 2 or 3. We present the images in a 4 by 4 grid 130 with two familiar images and the rest distractors. Three 4x4 tables of numbers are also presented 132, labeled 1, 2, and 3, and each table contains the numbers 1 to 16 in a permuted random order. The user looks up the number representing the location each previously seen picture in the table labeled by that picture's label, adds the numbers up and reports the answer, modulo 16. Two interactions of this scheme, given perfect recognition by the user, will reduce the change of imposture to 0.4%.

There are further schemes based broadly on Fig. 9, all taking advantage of a random element in each interaction, the random element being known to the system performing the certification, and a non-linear transformation, such as the use of the parity of the picture positions in the first example, or alternatively additional hidden information, for example the labels as learned in the second example hereinabove. They have the effect that the subject can train to higher accuracy of recognition on a smaller set of images and reuse those without need for retraining.

Reference is now made to Fig. 12 which is a screen shot illustrating an actual 4x5 grid of images used in a trial. In the following we discuss a protocol used in a further preferred embodiment that follows the principles outlined above of a challenge response protocol, where authentication is based on the user answering correctly a sequence of challenges posed by the computer. The challenges (or queries) are based on a shared secret between the computer and the user. The embodiment contributes a choice of secret - a long sequence of meaningful pictures, and a way of constructing the queries. The main advantage of the present embodiment is its feasibility and we demonstrate below that people can use the system correctly, and even enjoy it along the way. This relative user friendliness is accomplished at the cost of lower security; thus the protocol appears hard to break, but we cannot formally prove the level of difficulty in overcoming the arrangement. In the absence of formal analysis, we discuss probabilistic attacks of the protocol that demonstrate its power and limits.

Authentication protocol

The present embodiment comprises an authentication scheme as follows:

The computer assigns to each user two sets of pictures:

-- a set F of M familiar pictures;

-- a set B of N pictures, which includes the familiar set F .

Set B is common knowledge and may be fully or partially shared among different users; set F is arbitrarily selected for each user, and its composition is essentially the shared secret between the user and the computer.

During training in a secure location, the user is trained to recognize the set of familiar pictures F among pictures from the larger set B.

During authentication, the computer randomly challenges the user with the following query:

1. A set of n pictures is randomly selected from B.

2. The user is asked a simple multiple-choice question about the random set, which can be answered correctly only by someone who knows which pictures in the random set also belong to F .

3. The process is repeated k times; after each iteration, the computer computes the probability that the sequence of answers was generated by random guessing.

4. The computer stops and authenticates the user when the probability of guessing goes below a pre-fixed threshold. If this is not accomplished within a certain number of trials, the user is rejected.

In order for this scheme to work, we need to address a few questions:

How to construct a secret which is relatively easy for people to remember?

How to construct multiple-choice queries so that both conditions are met:

(i) users find the query manageable, and can answer correctly within a short time;

(ii) an eavesdropping adversary (our 'Eve') cannot learn the secret simply by recording successful authentication sessions.

How to conduct effective training?

How many sessions are safe? in other words, how many successful authentication sessions Eve must see, before she can successfully pose as the legitimate user?

ı ne nrst tnree questions are addressed in order in the following subsections, while the last question in addressed later on.

Shared secret

The shared secret between the human and the computer should be easy for people to remember, and hard for them to give away to other people. In addition, the secret should be relatively quick to memorize for most people, its memory trace should persist for a long time, and recognition should be relatively quick and precise. The first requirements is typically best served by automatic memory phenomena which require low awareness, such as procedural memory (a skill), perceptual learning or priming.

However, most of the candidate phenomena from perceptual learning and priming do not satisfy the second set of requirements: for some phenomena the length of training is unreasonably long, for others memory persistence is too short-term, while for others still the measurement of secret knowledge is ``unacceptably" tedious.

We therefore chose the explicit recognition of memory items. Pictures emerged as the most promising modality, since it appears that people can remember a vast amount of pictures for a relatively long time after only a short exposure for a few seconds. Pictures are also hard to describe (or be given away) to other people. We have also demonstrated the feasibility of other memory modalities, such as pseudo words, as discussed above, which can be used when pictures are not appropriate.

In the present embodiment, each user is assigned a public set of N = 240 pictures B, and a secret subset of M = 60 familiar pictures F # B. The length of the secret, or the total number of all possible different subsets F , is $\binom{240}{60} \approx 2.4.10^{57} \approx 2^{190}$. Thus the length of the secret is roughly 190 bits, which is long enough according to the rules of thumb recognized in the field. At the same time, people can easily memorize 60 pictures for future recognition, and even memorizing 240 pictures can be done in relatively short time.

Query construction

The query is constructed with two opposite goals in mind: On the one hand, the query should be easy for users to compute unassisted, and cannot therefore include complex mathematical operations. On the other hand, the correct answer to the query should not reveal to an outsider too much about the shared secret. By this we mean the following: an adversary, who observed L correct answers to any set of L random queries, will not be able to compute the secret in reasonable time with reasonable computational power. We discuss next the query used in the embodiment and argue that it is relatively easy for people to use. The protocol's security is discussed hereinbelow. Specifically in our protocol, a query is asked with respect to a panel of n = 20 pictures randomly selected from the set B. The pictures are shown on a regular grid, including 4 rows and 5 columns, as shown in Fig. 12. Each picture is assigned a random bit (0 or 1), which is shown next to it. Preferably, the bits are balanced, with 10 random pictures assigned 0 and the rest assigned 1. The user is instructed to scan the panel in order: from left to right, one row after another.

We studies two variants of possible binary queries:

1. The user is asked to identify the first familiar picture (the first picture from subset F), and the last familiar picture. She should then compare the associated bits, and answer whether they are the same or different.

2. The user is asked to identify the first, second and last familiar pictures (from subset F). She should then compare the 3 associated bits, and answer whether their majority is 0 or 1. Parenthetically, the user may separately be instructed what to do when there are less than 2 familiar pictures in the first case, or less than three in the second case.

The second variant above of the query is a little more difficult for users to compute, but significantly harder for eavesdroppers to use for secret discovery.

In our implementations we used two more query variants: these are identical to the two above, with one difference:

the bit attached to each picture is not shown on the screen (visible to Eve), but is effectively presented in such a way that only the legitimate user can see. Specifically, during training a subject also learns a simple cue attached to each picture. During authentication, the existence of the cue is treated as bit value 1, while the absence of the cue is treated as bit value 0. The cue actually used is described hereinbelow although the skilled person will appreciate that other variations are possible.

in connection with the present embodiment we have described so far four specific query variants, but clearly other variants could be used for which the problem of key discovery may be possibly harder. For example, a non linear function computed using all the familiar pictures in the panel will be hard to use by an eavesdropping adversary. We chose to use only 2 or 3 familiar pictures in order to make the task easier for average users with high reliability.

Training and fortifying user's memory

In the training (or familiarization) phase, 100 in Fig. 9, the user is familiarized with the subset of pictures F only. In each training session all the pictures from F are shown once in random order.

Each picture is shown for a few seconds, after which the user is asked to memorize the picture for a second or so, and then answer a multiple choice question which depends on the details of the memorized picture. A training session lasts on average 15 minutes. Three training sessions are given, typically on three consecutive days.

We experimented with two kinds of memory questions, as illustrated in Figs. 13a and 13b, to which reference is now made. Fig. 13a shows a change detection paradigm: an original picture 140 is shown for 400 ms, followed by a mask 144 which is shown for 100 ms, followed by a modified image 142 shown for 400 ms. The first and last pictures differ in a single detail: one ball has disappeared in the transition between the pictures. Fig. 13b is an example from the training session: and shows a multiple choice graphic query between 4 low contrast images.

In Fig. 13b, the user is shown four low contrast pictures in a 2 2 ' grid. This set of four pictures includes an original picture, and three pictures which are obtained by geometrical ransformations of the original picture:

reflection around the X- axis,

reflection around the Y - axis, and

rotation by 180°.

The pictures are shown in random order, and the user must click on the original (not transformed) picture. If the user chooses wrongly, the contrast is increased, the pictures are rearranged, and the user is asked to choose again. This is repeated until the user nnds the correct picture. 'I'he user receives points for each correct selection: the lower the contrast, the more point she gets.

The second kind of memory question used the change detection paradigm to solidify the memorization of pictures. This variant has the advantage that each memorized picture is also connected in the user's mind with an independent arbitrary cue. Specifically, during the construction of the database, each picture was assigned a pair, another picture which is similar to it in almost everything but a small detail, as in the two images of Fig. 13a.

In the initial presentation of each picture 140, it is shown flickering with its pair 142, with a short blank mask 144 in between. Next, the user is shown 3 flickering pairs side by side: the original picture flickering with itself, the modified picture flickering with itself, and the original picture flickering with its modified pair. Only in the third condition is there a change between the two flickering pictures. The 3 flickering pairs are randomly arranged, and the user is asked to click on the pair where there is a change between the flickering images. If the user chooses wrongly, he is shown the original image flickering with its pair again, but without the interleaving mask. This makes the detection of the change immediate. The multiple choice selection is then repeated until the user gets it right.

The training scheme described so far seems to suffer from a major flaw. Indeed training only on pictures from F makes those pictures familiar to the user, who will experience a feeling of 'deja vu' when seeing one of these pictures in a later authentication session. However, recall that during authentication each query involves a random subset of 20 pictures from B, and therefore pretty soon the user will have seen a large fraction of the pictures in the whole set B. Thus the user may develop a sense of familiarity with all the pictures in B, and pretty soon will not be able to reliably distinguish F pictures from B pictures. This could make it harder and harder for the user to answer authentication queries correctly.

Although this seems like a major problem, in our experiments it proved to be insignificant, as discussed below. Possibly because users are familiarized with pictures in F and pictures in {B \ F} in different ways and at different times, it appears that most manage to keep the sets separate in their mind.

Such a cue is the one we have used in the two cued query variants described in the previous description. The cue is implemented as follows: When presented with a query, the user has an option to click a ``flicker" button, whereby all the 20 pictures in the panel are flickered once with their matching pairs. The user can use this option to focus on one picture at a time. A cue exists when during flicker the picture is flickered with its pair, showing the trained change. A cue does not exist when during flicker the picture is flickered with another pair, showing an untrained change, which the trained user cannot see.

User study and protocol feasibility

Three subjects, BS, AS, and OK, were trained and tested as described above, and the results are presented in Fig. 14, which is a graph of success rate against number of days for each of the three subjects. Fig. 14 shows the results of the user study with the three subjects, showing the average success rate in answering a set of approximately 20 queries, as a function of the time passed since the last training. All subjects performed perfectly in the first 10 days, and therefore the curves were artificially shifted for better visualization.

Subjects did very well in the period studied, which as 6 weeks for one subject, 3 weeks for the second, and 2.5 weeks for the third. Subjects AS and OK were tested with the four query variants, and showed similar performance with all four of them.

These results suggest that the protocol may be practical for user authentication. With a safety threshold of 1 in a million (i.e., accept user only if the chance of guessing is less than 10 -6), a user will be authenticated after 20 queries when all his answers are correct, which means that the session will take approximately 1.5 minutes to conclude (estimating 5 seconds per query).

How safe is the protocol from eavesdropping adversaries

The problem that Eve needs to solve in order to discover a user's key is the following: given L queries and their 1 bit answer, and given the shared set B of N pictures, what are the M pictures that define the user's secret subset F? Eve doesn't know that the user's answer is always correct. In addition, L can't be too large: if each query takes roughly 5 seconds to answer, it seems fairly safe to assume that L is not larger than 100,000, and rarely larger than 10,000.

In the following discussion, we discuss probabilistic attacks against the protocol. Eve's problem appears hard, though we do not include a proof that it formally is hard. The discussion also justifies our choice of specific queries.

Since the answer of the user ultimately depends only on two or three pictures in each panel, we consider all the pairs or all the triplets in the set B. For N = 240, the number of all pairs is 28, 680, and the number of all triplets is 2, 275, 280. Given a query and the user's answer, we come up with a voting scheme that will give on average higher scores to pairs and triplets in F , as compared to the remaining pairs or triplets in B. After enough observations (large L) we estimate reliably which pairs or triplets consistently get higher votes, from which we derive an estimate of F .

Table 1 below shows the results of the probabilistic attack we have implemented against two of the query variants - those relying on only the first and last familiar pictures in the query panel.

Build a table of size N 'N (initialized to 0), to represent all pairs of images.

For each query:

Check if the first and last pictures in the panel are consistent with the user's answer. If not, eliminate this pair completely since it cannot be in F (this is an absolute final veto vote).

For each pair in the panel (total of 2 n pairs), add a vote whose value is inversely proportional to the distance between the pictures in the panel (where distance varies from 1 to n - 1). The sign of the vote reflects whether the pair matches the correct answer (positive sign) or not (negative sign).

After L queries, estimate F using the following greedy algorithms:

-- select the picture with the largest (remaining) marginal vote; 3

-- eliminate all points which are linked to it as an impossible pair, and strengthen the total value of all the pairs which include the selected point;

-- repeat M times.

Simulate a user who uses the estimated F to enter the system, and measure his success rate by the average number of his correct answers.

|B|	L	p	std
240	1000	0.52	0.03
240	2000	0.63	0.16
240	3000	0.72	0.2
300	2000	0.54	0.06

|B|	L	p	std
240	2000	0.52	0.03
240	3500	0.57	0.07
240	5000	0.69	0.07
300	5000	0.58	0.08

Table 1: Upper set shows results for the first query variant: the user reports whether the bits attached to the first and last familiar pictures are the same or not. Lower set: results for the third query variant: the user reports whether the visibility of the trained change in the first and last familiar pictures is the same. The columns in each table are as follows: size of shared set B, number or simulation runs, mean and standard deviation for the success rate of an imposter over many simulated attacks.

We may conclude from the results in Table 1 that the two query variants which rely only on two familiar pictures in a query panel are only moderately secure. Not surprisingly the more complicated variant which uses the hidden cue is more secure. But even with this second variant, the user cannot use this scheme for more than 250 entries, assuming the user performs perfectly and gets authenticated after 20 queries to a total of 5000 queries.

Although we did not simulate attacks against the two query variants which rely on three familiar pictures, clearly a similar attack to the one described above will require many more observed sessions in order to succeed. If our simulated attack reflects in any way the best possible attack against the protocol, it would appear that the variants which use three familiar pictures are sufficiently secure for any feasible number or entries. Fig. 15 illustrates different imposter acceptance rates against different parameters of the query, and taking into account days since training of the subject. It is apparent from the graph that beyond around 30 days after training the genuine user is close to indistinguishable from the imposter.

Although the invention has been described in conjunction with specific embodiments thereof, it is evident that many alternatives, modifications and variations will be apparent to those skilled in the art. Accordingly, it is intended to embrace all such alternatives, modifications and variations that fall within the spirit and broad scope of the appended claims. All publications, patents and patent applications mentioned in this specification are herein incorporated in their entirety by reference into the specification, to the same extent as if each individual publication, patent and patent application was specifically and individually indicated to be incorporated herein by reference. In addition, citation or identification of any reference in this application shall not be construed as an admission that such reference is available as prior art to the present invention.

кeierences:

Reber, A. S. (1967). Implicit learning of artificial grammars. Journal of Verbal Learning and Verbal Behavior, 6, 855-863.

A. J. Parkin (2000). Essential cognitive psychology. Psychology Press LTD.

Perruchet, P. and Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? Journal of Experimental Psychology: General, 119, 264-275.

E. Tulving, D. L. Schacter, H. A. Stark (1982). Priming effects in word-gragment completion are independent of recognition memory. Journal of Experimental PSychology: Learning, Memory & Cognition, 8(4):336-342.

Cave, B. C. Very long-lasting priming in picture naming. Psychol. Sci. 8, 322-325 (1997).

G. Musen and A. Treisman (1990). Implicit and Explicit Memory for Visual Patterns. J Exp Psychol Learn Mem Cogn, 16(1):127-37.

Rensink, R. A., O'Regan, J. K., and Clark, J. (1997). To see or not to see: the need for attention to perceive changes in scenes. Psychological Science, 8(5), 368-373.

Long, G. and Olszweski, D. (1999). To reverse or not to reverse: When is an ambiguous figure not ambiguous? American Journal of Psychology, 112, 41-56.

R. N. Shepard (1967). Recognition memory for words, sentences, and pictures. J Verb Learn Verb Behav, 6:156-163.

A. Karni and D. Sagi (1993). The time course of learning a visual skill. Nature, 365, 250-252.

R. Dhamija and A. Perrig (2000). Déjà vu: A user study using images for authentication. In *Proceedings of the 9th USENIX Security Symposium,* 2000.

N. J. Hopper and M. Blum (2000). A secure human-computer authentication scheme, preprint CMU-CS-00-139.

T. Matsumoto (1996). Human-computer cryptography: an attempt. In *ACM Conference on Computer and Communications Security,* pp. 68-75, 1996.

T. Matsumoto (1991). Human identification through insecure channel. In *Theory and Application of Cryptographic Techniques,* pp.409 – 421, 1991.

Microsoft. A press report is given at http://research.microsoft.com/displayArticle.aspx?id=417

Password. http://www.atstake.com/research/lc

R. Dhamija and A. Perrig. Dj vu: A user study using images for authentication. In In Proc. 9th
USENIX Security Symposium, 2000.

N. J. Hopper and M. Blum. Secure human identification protocols. In In Proc. Advances in
Cryptology, pages 52--66, 2001.

S. Li and H.-Y. Shum. Sechci: Secure human-computer identification (interface) systems
against peeping attacks, 2003.

A. Menezes, P.C. van Oorschot, and S.A. Vanstone. handbook of Applied Cryptography. CRC

Press, 2nd edition, 1996.

M. Naor and B. Pinkas. Visual authentication and identification. In In Proc. Advances in Cryptology, pages 322--336, 1997.

R.A. Rensink, J.K. O'Regan, and J.J. Clark. On the failure to detect changes in scenes across brief interruptions. Visual Cognition, 7:127--145, 2000.

A. Salaso, R. M. Shiffrin, and T. C. Feustel. Building permanent memory codes: Codification and repetition effects in word identification. J Exp Psyc: General, 114(1):50--77, 1985.

L. Standing, J. Conezio, and R. N. Haber. Perception and memory for pictures: single trial learning of 2500 visual stimuli. Psychol. Sci., 19:73--74, 1970.

D. Weinshall and S. Kirkpatrick. Passwords you'll never forget, but can't recall. In In Proc. Conf. on Computer Human Interfaces, 2004.

WHAT IS CLAIMED IS:

1. A method for providing a security function with a user, comprising:

imprinting the user with at least one cryptographic primitive determined from a sensory mechanism; and

at least one of authorizing, identifying or authenticating the user according to an ability to recall said at least one cryptographic primitive.

2. The method of claim 1, wherein said at least one of authorizing, identifying and authenticating comprises measuring user knowledge of one fact of said cryptographic primitive via a query presented to said user.

3. The method of claim 1, wherein said at least one of authorizing, identifying and authenticating comprises measuring user knowledge of two facts of said cryptographic primitive via a query presented to said user.

4. The method of claim 3, wherein said imprinting said cryptographic primitive comprises learning an image, and one of said facts is that a given image is a learnt image and a second of said facts is associated with a position of said learnt image in a grid.

5. The method of claim 3, wherein said imprinting said cryptographic primitive comprises learning an image and one of said facts is that a given image is a learnt image and a second of said facts is associated with learnt changes in said image.

6. The method of claim 1, comprising repeating said measuring for different cryptographic primitives for a predetermined number of times.

7. The method of claim 1, comprising repeating said measuring for different cryptographic primitives until a predetermined probability threshold of correct authentication is reached.

8. The method of claim 1, wherein said imprinting comprises implicit learning by the user.

9. The method of claim 8, wherein said at least one cryptographic primitive is used to encrypt a message according to a one-way function.

10. The method of claim 8, wherein a one-time pad comprises said at least one cryptographic primitive.

11. The method of claim 8, wherein a near-zero knowledge function comprises said at least one cryptographic primitive.

12. The method of claim 8, wherein said sensory mechanism comprises vision, such that said at least one cryptographic primitive comprises recognizing an image.

13. The method of claim 12, wherein said recognizing said image comprises:

training the user on a plurality of trained images; and

testing the user on a combination of a trained image with at least one distractor image.

14. The method of claim 13, wherein said at least one distractor image comprises a plurality of distractor images.

15. The method of claim 13, wherein said testing comprises:

selecting a plurality of different trained images by the user in sequence, said sequence providing said cryptographic primitive for determining said at least one of authorizing, identifying or authenticating the user.

16. A method for authenticating, authorizing or identifying a user, comprising:

training the user with information through a sensory mechanism; and

determining accurate recall of said information to authenticate, authorize or identify the user.

17. A method for a one-way function for authenticating, authorizing or identifying a user, comprising:

imprinting the user with a cryptographic primitive; and

testing said imprinting with at least a similar or identical cryptographic primitive to authenticate, authorize or identify the user.

18. The method of claim 17, wherein said cryptographic primitive is derived from input according to a sensory mechanism.

19. The method of claim 18, wherein said input comprises at least one image and said sensory mechanism is visual.

20. The method of claim 18, wherein said input comprises at least one pseudoword and said sensory mechanism is verbal.

21. The method of claim 18, wherein said sensory mechanism is selected from the group consisting of tactile, olfactory, audible and taste.

22. The method of claim 17, wherein said testing comprises determining whether the user is capable of discriminating between an imprinted cryptographic primitive and a non-imprinted cryptographic primitive.

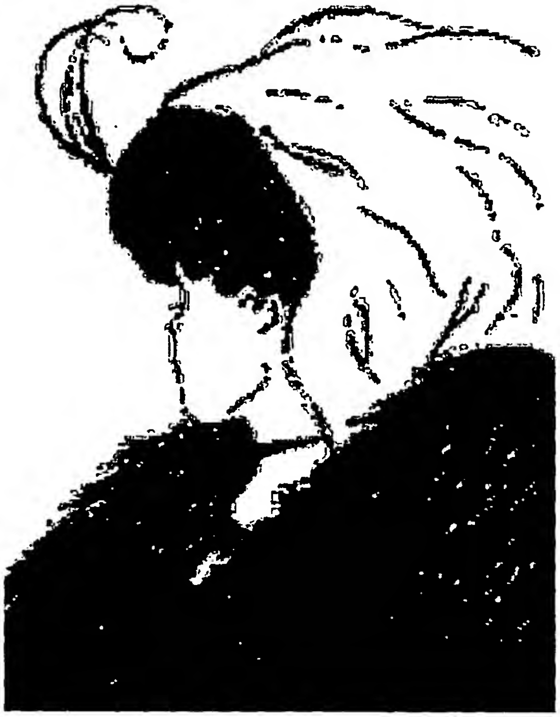

Fig. 1

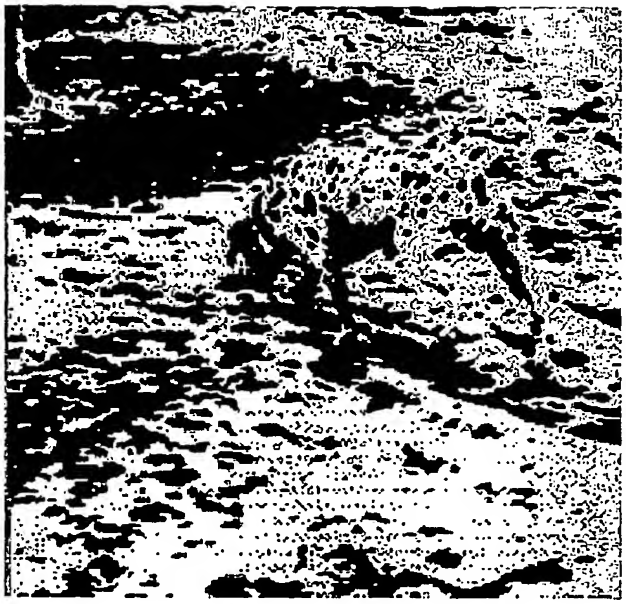

Fig. 2

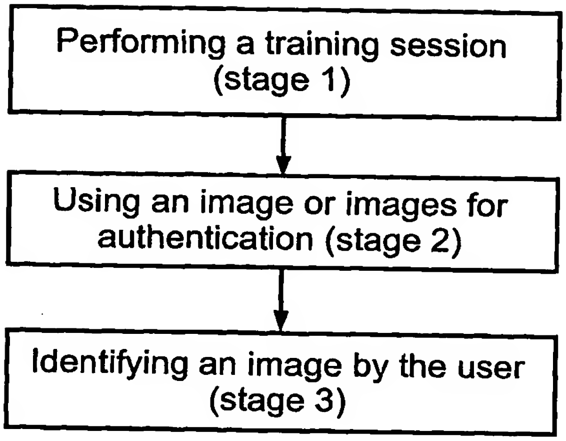
Performing a training session (stage 1)
Using an image or images for authentication (stage 2)
Identifying an image by the user (stage 3)

Fig. 3

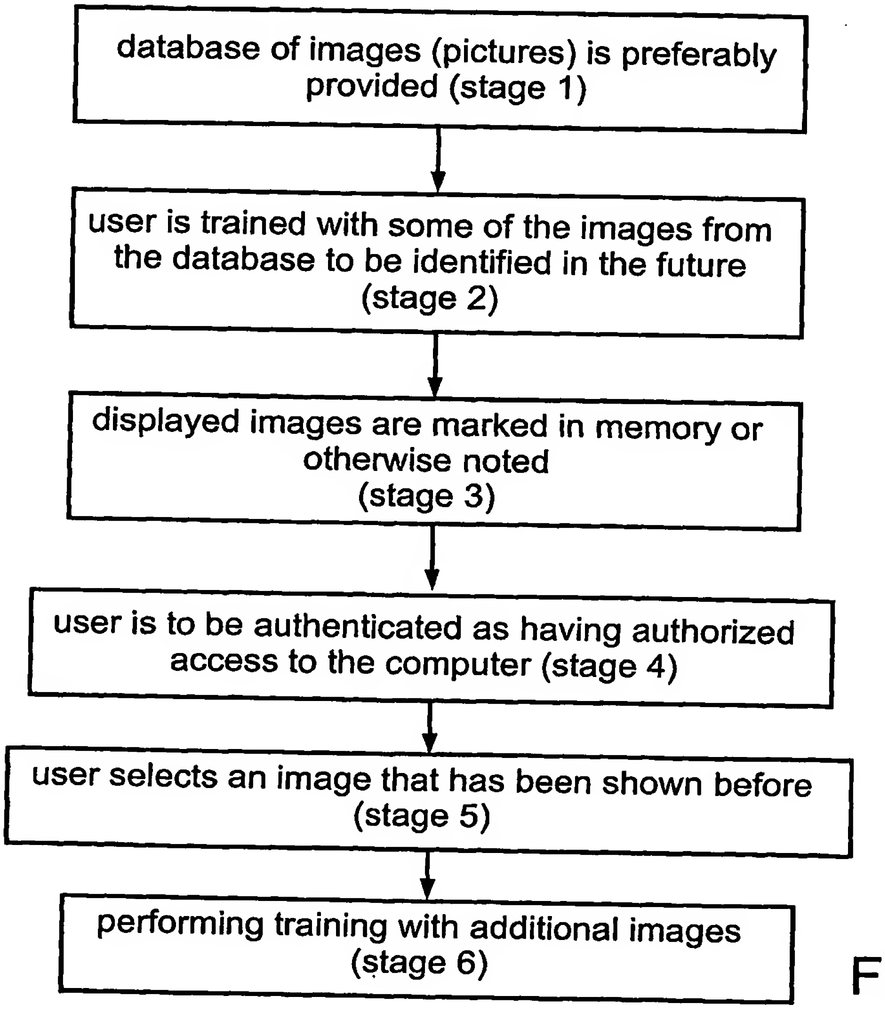
database of images (pictures) is preferably provided (stage 1)
user is trained with some of the images from the database to be identified in the future (stage 2)
displayed images are marked in memory or otherwise noted (stage 3)
user is to be authenticated as having authorized access to the computer (stage 4)
user selects an image that has been shown before (stage 5)
performing training with additional images (stage 6)

Fig. 4

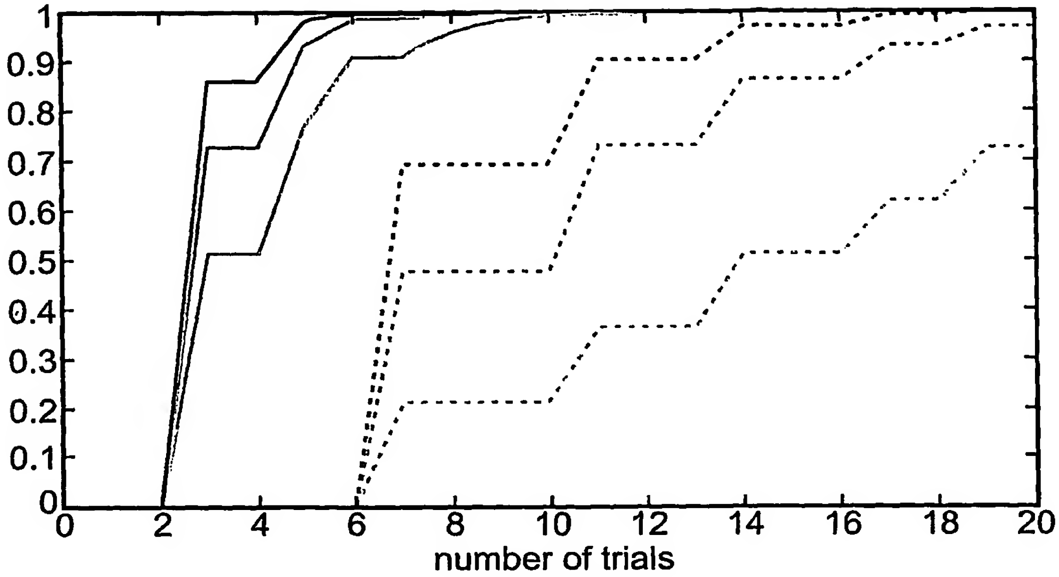
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
12
14
16
18
20
number of trials

Fig. 5a

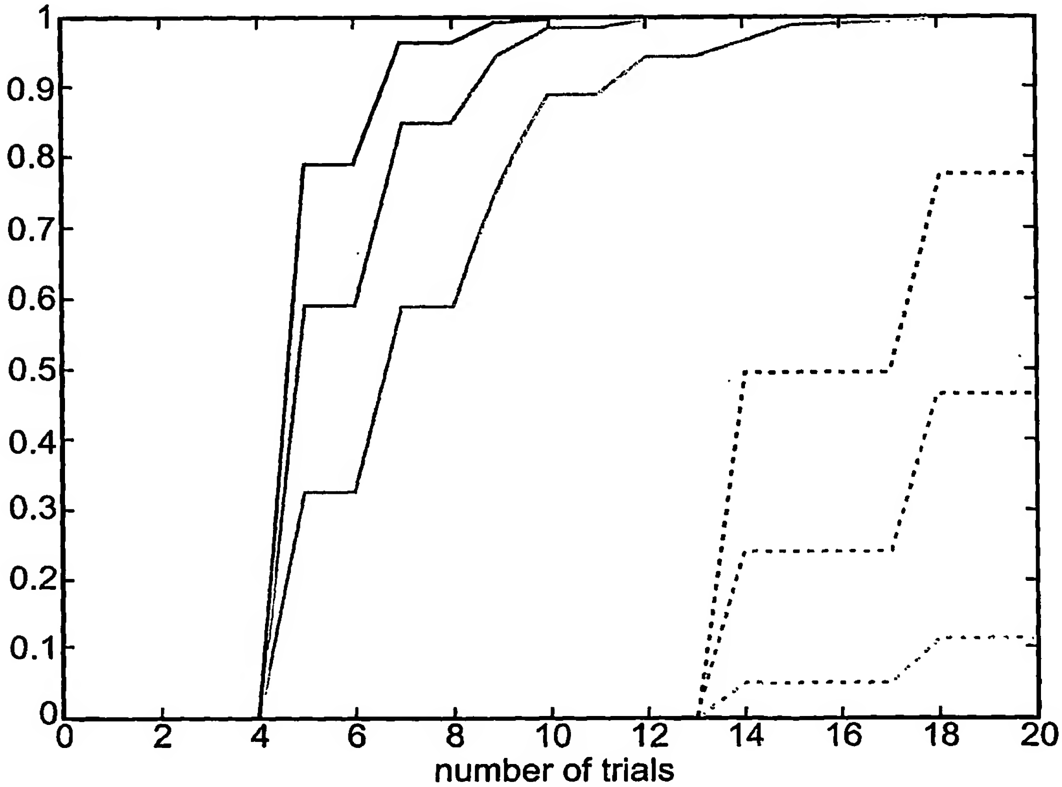
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
12
14
16
18
20
number of trials

Fig. 5b

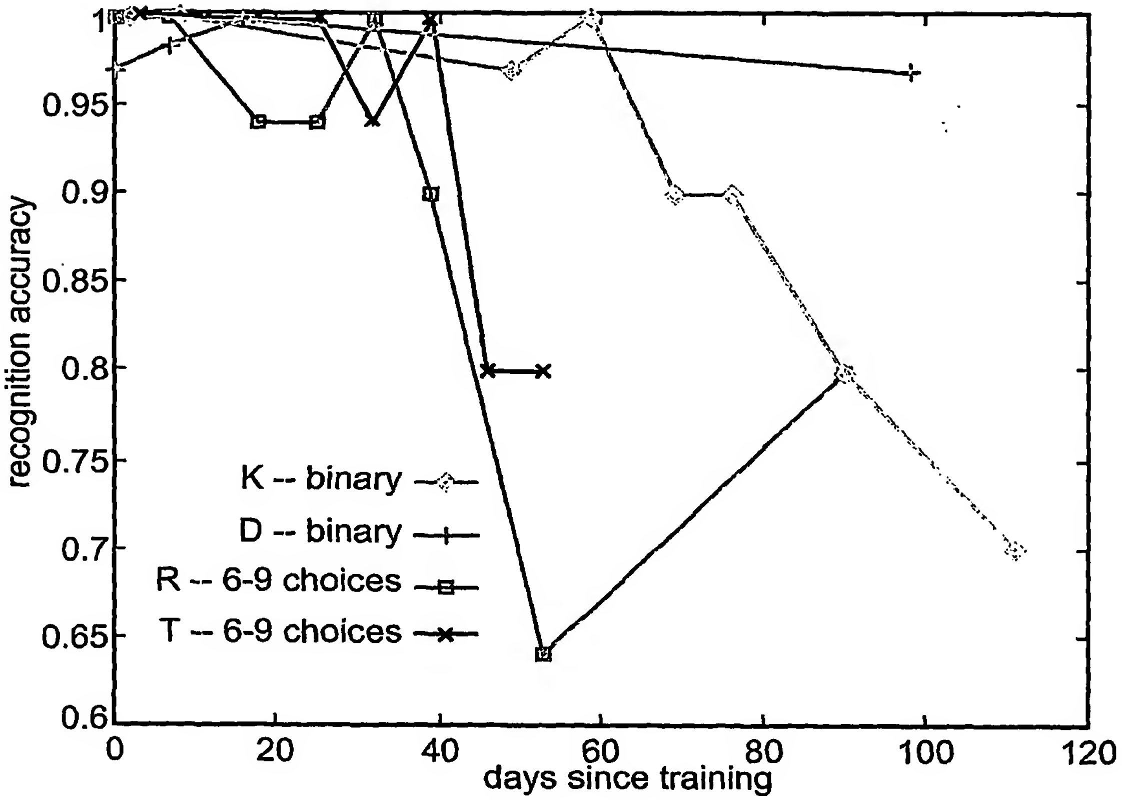
recognition accuracy
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0
20
40
60
80
100
120
days since training
K -- binary
D -- binary
R -- 6-9 choices
T -- 6-9 choices

Fig. 6

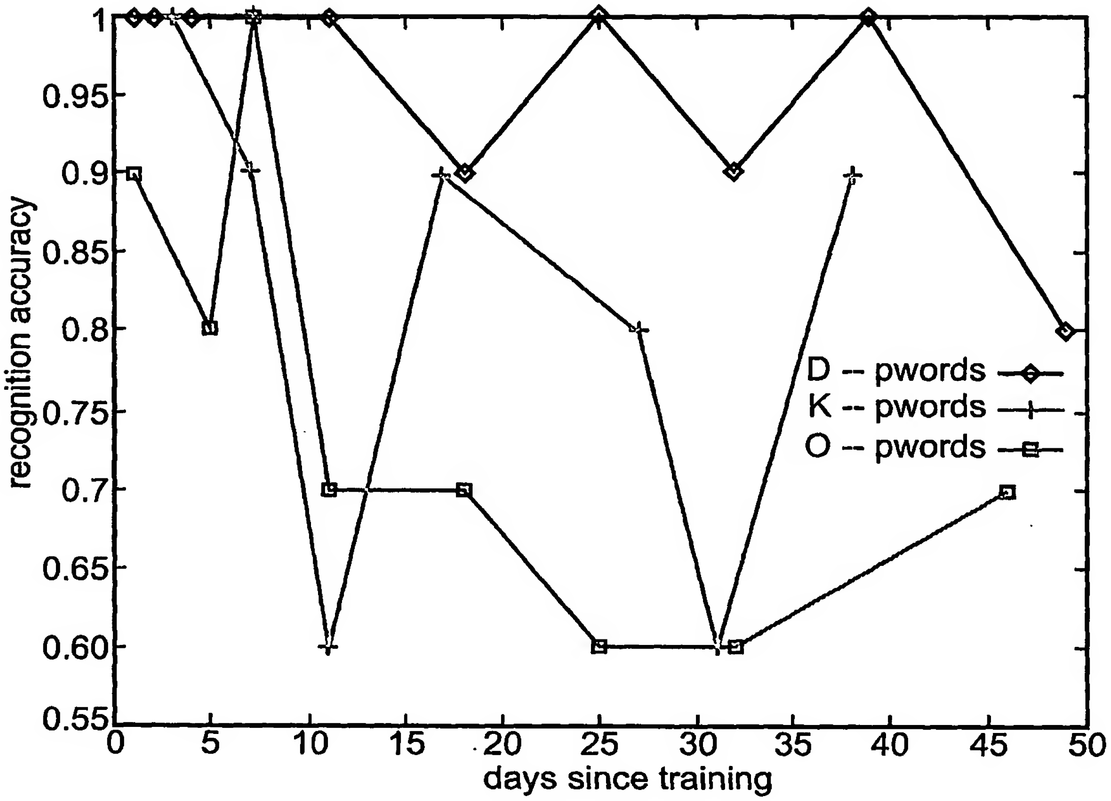
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.60
0.55
recognition accuracy
0
5
10
15
20
25
30
35
40
45
50
days since training
D -- pwords
K -- pwords
O -- pwords

Fig. 7

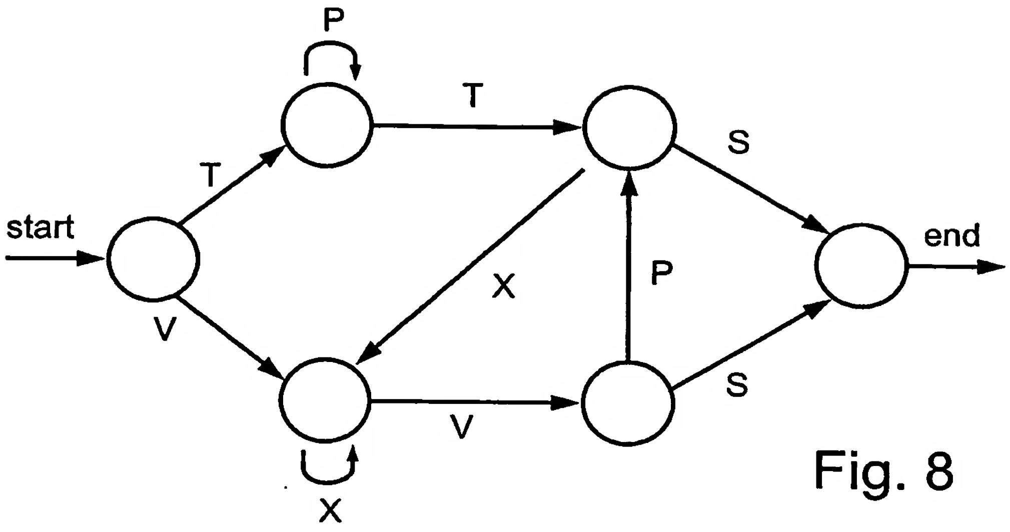
P
T
S
T
start
end
X
P
V
S
V
X

Fig. 8

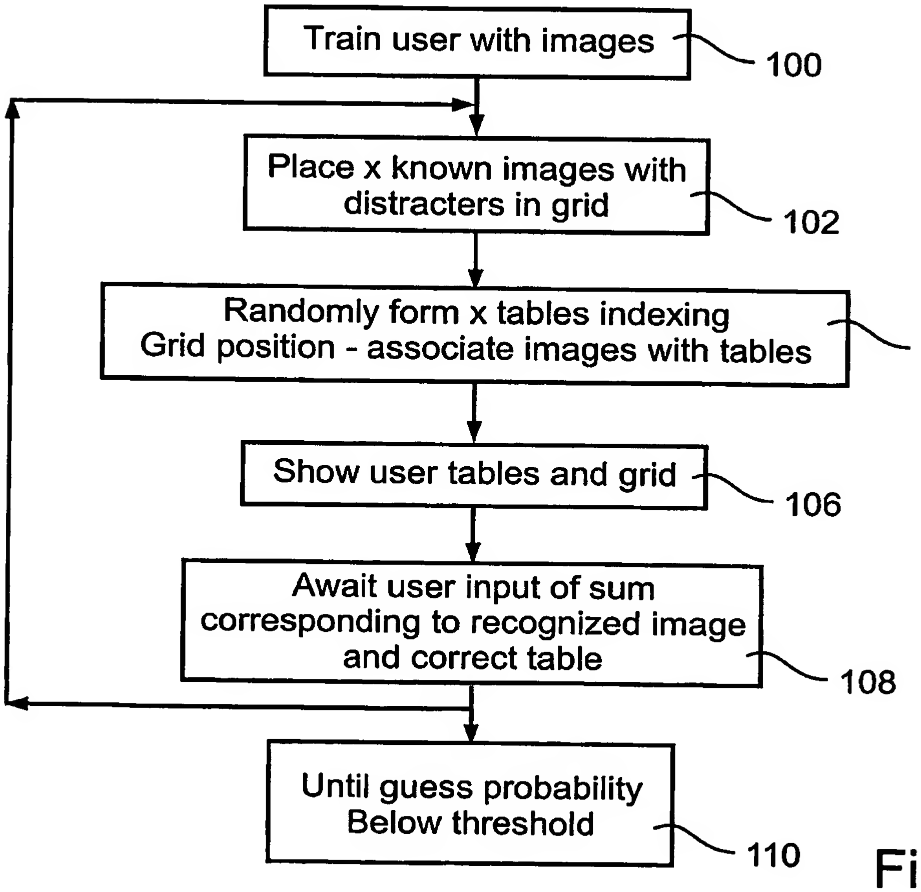
Train user with images
100
Place x known images with distracters in grid
102
Randomly form x tables indexing Grid position - associate images with tables
104
Show user tables and grid
106
Await user input of sum corresponding to recognized image and correct table
108
Until guess probability Below threshold
110

Fig. 9

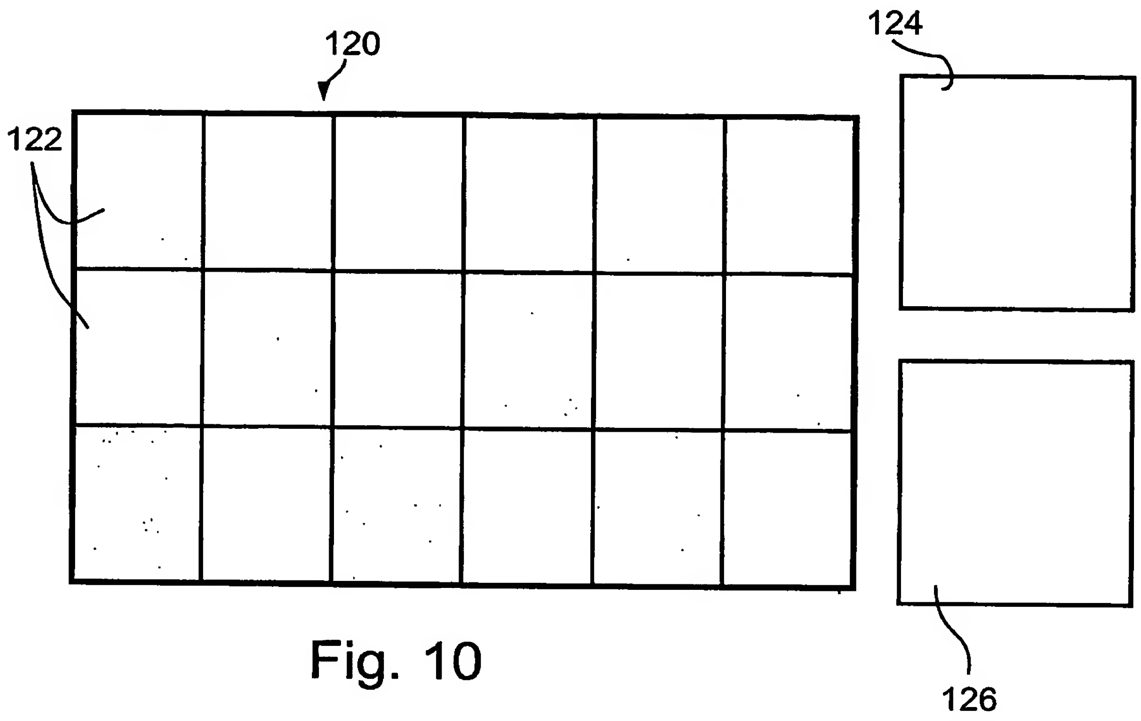

Fig. 10

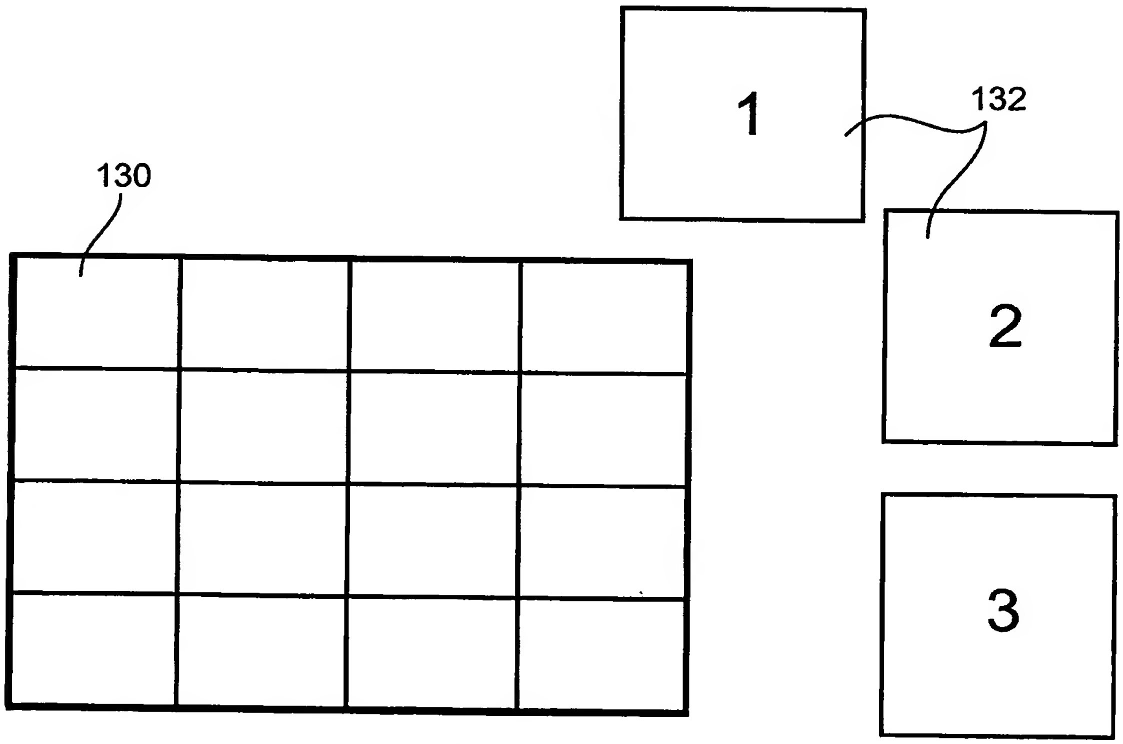

Fig. 11

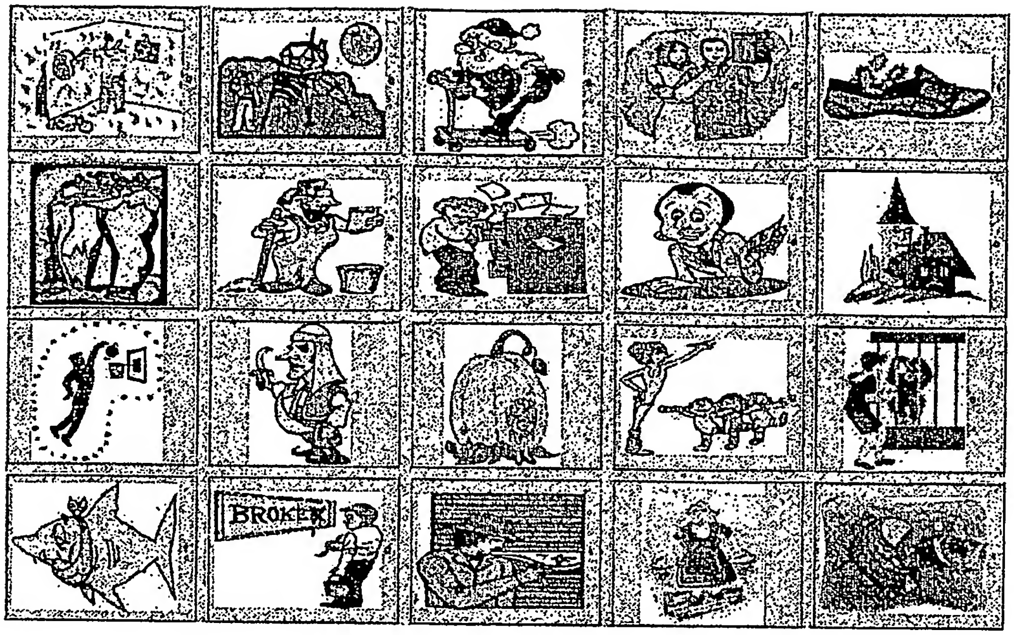

Fig. 12

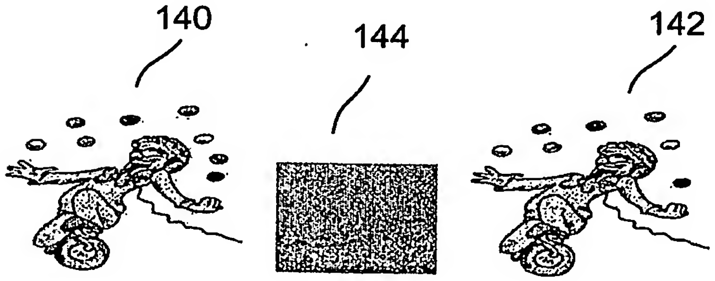

Fig. 13a

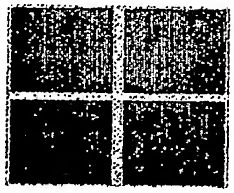

Fig. 13b

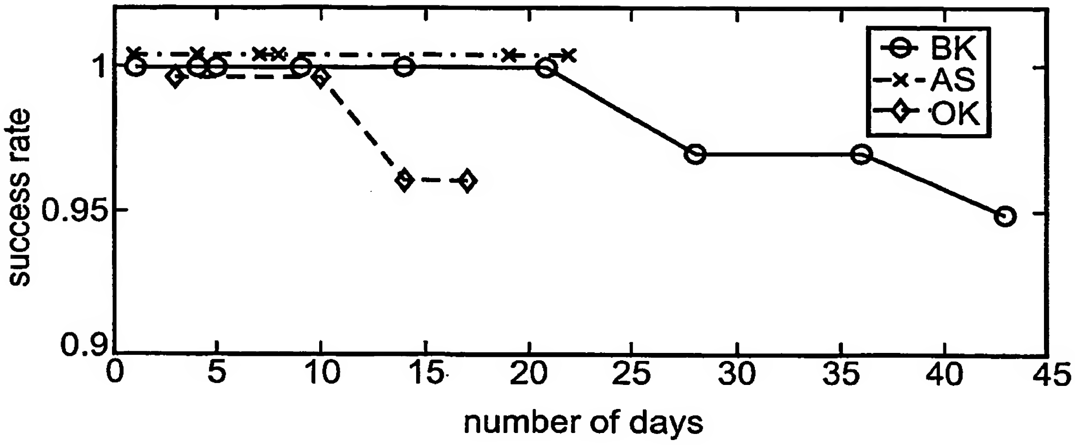
success rate
1
0.95
0.9
0
5
10
15
20
25
30
35
40
45
number of days
BK
AS
OK

Fig. 14

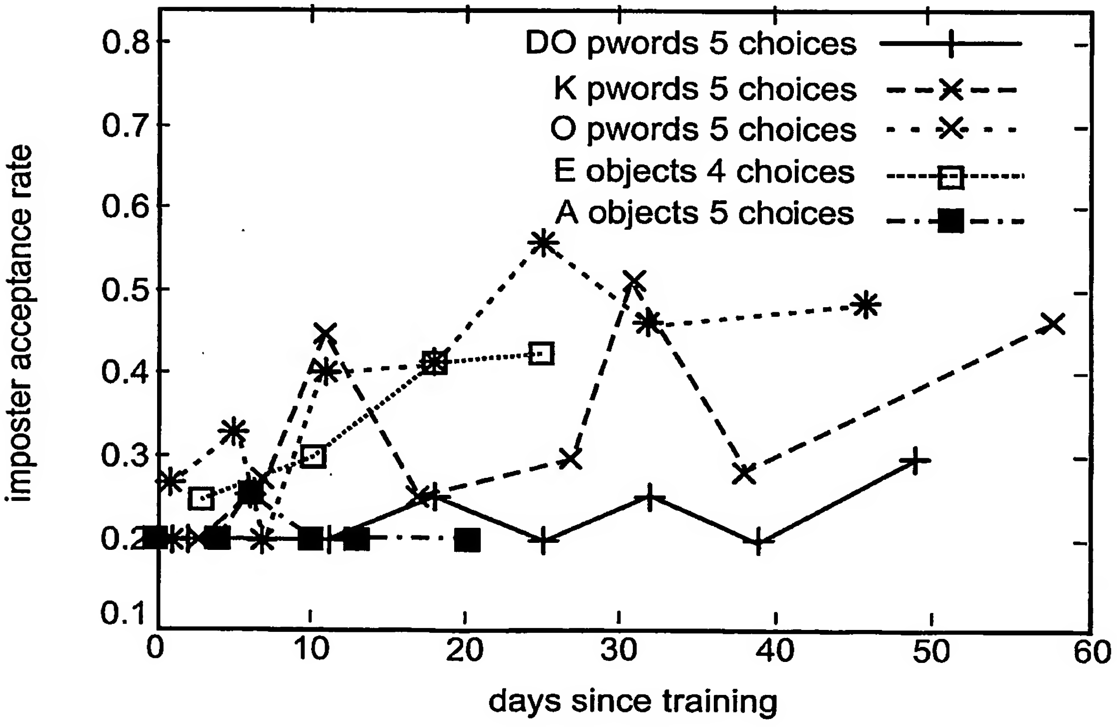
imposter acceptance rate
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
10
20
30
40
50
60
days since training
DO pwords 5 choices
K pwords 5 choices
O pwords 5 choices
E objects 4 choices
A objects 5 choices

Fig. 15